

Dr Agnieszka MAZUR-DUDZIŃSKA
Politechnika Łódzka, Katedra Zarządzania

SZTUCZNE SIECI NEURONOWE W MODELOWANIU ZJAWISK ZACHODZĄCYCH NA RYNKU NIERUCHOMOŚCI

Streszczenie: Celem referatu jest zastosowanie sztucznych sieci neuronowych do analizy zjawisk zachodzących na rynku nieruchomości. Szczególnie istotnym zagadnieniem związanym z tym rynkiem jest proces wyceny nieruchomości, dlatego podjęto próbę zastosowania sieci neuronowych do przewidywania cen nieruchomości. Problem ten został potraktowany jako zagadnienie klasyfikacji nieruchomości do jednej z grup cenowych. Badania zostały przeprowadzone na podstawie danych empirycznych pochodzących z lokalnego rynku mieszkaniowego. Do klasyfikacji obiektów zastosowano jednokierunkowe sieci wielowarstwowe.

APPLICATION OF THE ARTIFICIAL NEURAL NETWORKS TO THE REAL ESTATE MARKET ANALYSIS

Abstract: The aim of the paper is the application of the artificial neural networks to analysis the economic phenomena which take place in the real estate market. One of the most important issue related to the market is the estimation of real estate price. That is the reason why it was decided to attempt of neural networks in order to foresee real estate prices. The investigation is provided on the basis of the empirical data from local apartment market. The apartments were classified to one of the price groups. For classification of objects feedforwards networks were used.

Słowa kluczowe: sztuczne sieci neuronowe, metody klasyfikacji, rynek nieruchomości, wycena nieruchomości

Keywords: artificial neural networks, classification methods, real estate market, estimation of real estate price

1. WPROWADZENIE

W dobie nieustannie rosnącej ilości informacji, a co za tym idzie coraz większego zapotrzebowania na systemy ich przetwarzania, przy jednoczesnym wysokim rozwoju możliwości technicznych, dużą popularność zdobyły tzw. metody sztucznej inteligencji. Do grupy tych metod zalicza się sztuczne sieci neuronowe, które są nowoczesnymi systemami obliczeniowymi przetwarzającymi informacje w sposób wzorowany na zasadach funkcjonowania układu nerwowego człowieka.

Duża przydatność sieci neuronowych w rozwiązywaniu wielu złożonych problemów wynika z ich zdolności równoległego przetwarzania informacji, możliwości uczenia, a także uogólniania zdobytej wiedzy. Poza tym cechuje je:

- zdolność przetwarzania danych, które są niekompletne,
- możliwość wytwarzania rezultatów przybliżonych,
- szybkie i efektywne przetwarzanie dużej liczby danych,
- przetwarzanie równoległe, rozproszone,
- skojarzeniowy dostęp do informacji zawartej w sieci,
- informacja rozproszona (zawarta w połączeniach),

- nadmiarowość struktury i związana z tym duża tolerancja na błędy i uszkodzenia,
- możliwość przetwarzania informacji rozmytych, chaotycznych, a nawet sprzecznych [1].

Sztuczne sieci neuronowe okazały się bardzo wygodnym i przydatnym narzędziem nie tylko do rozwiązywania problemów z zakresu nauk technicznych, przyrodniczych czy medycznych, ale mają również bardzo szerokie zastosowanie w naukach społeczno-ekonomicznych. Jako przykłady wykorzystania modeli neuronowych w tym obszarze można wymienić badanie zdolności kredytowej przedsiębiorstw, analizę rynków kapitałowych, analizy marketingowe, prognozowanie wskaźników makroekonomicznych, finansowych czy prognozowanie zapotrzebowania na energię elektryczną. Z powodzeniem można je też wykorzystywać do modelowania zjawisk zachodzących na rynku nieruchomości. Celem referatu jest przedstawienie wyników zastosowania sieci neuronowych do przewidywania cen nieruchomości. Badania zostały przeprowadzone na podstawie danych empirycznych pochodzących z lokalnego (łódzkiego) rynku mieszkaniowego. Problem szacowania ceny nieruchomości został potraktowany jako zagadnienie klasyfikacji nieruchomości do jednej z (arbitralnie ustalonych) grup cenowych. Wszystkie eksperymenty numeryczne zostały przeprowadzone przy wykorzystaniu programu komputerowego *STATISTICA Neural Networks* v. 4.0.

2. RYNEK I WYCENA NIERUCHOMOŚCI

Jedną z definicji określa rynek nieruchomości jako ogół warunków, w których odbywa się transfer praw do nieruchomości oraz są zawierane umowy stwarzające wzajemne prawa i obowiązki, połączone z władaniem nieruchomości [2]. Rynek nieruchomości jest rynkiem posiadającym specyficzne cechy, podobnie jak sama nieruchomość. W odróżnieniu od innych rodzajów dóbr inwestycyjnych, nieruchomości charakteryzują między innymi [2, 3]: stałość w miejscu, trwałość, lokalny charakter, duża kapitałochłonność oraz mała płynność. To powoduje, że rynek nieruchomości ma szczególny charakter. Cechuje go między innymi wysoki poziom niedoskonałości, mała elastyczność i niska efektywność. Zarówno popyt, jak i podaż nieruchomości bardzo wolno reagują na ceny. Podaż w krótkim okresie czasu jest stała, a na jej kształtowanie się, podobnie jak na kształtowanie się popytu, ma wpływ wiele czynników pozacenowych (takich jak moda, tradycja czy sąsiedztwo). Ponadto, rynek nieruchomości wyróżnia specyfika transakcji rynkowych, które cechują między innymi wysoka wartość i koszty, konieczność obsługi prawnej oraz nierówny dostęp obu stron do informacji o przedmiocie transakcji [4].

Jednym z podstawowych zagadnień związanych z rynkiem nieruchomości jest określanie wartości nieruchomości. Do celów wyceny zalicza się przede wszystkim transakcje rynkowe (kupna-sprzedaży), zabezpieczenie kredytu, ubezpieczanie nieruchomości, zmianę przeznaczenia nieruchomości, transakcje pozarynkowe, np. darowizny, zmiana właściciela w wyniku postępowania spadkowego. Wycena jest niezbędna w przypadku aktualizacji opłat z tytułu wieczystego użytkowania gruntu, a także planowania przestrzennego oraz powszechnej taksacji na potrzeby naliczania podatku katastralnego. Proces wyceny jest ujmowany w procedury postępowania składające się z wyboru podejścia, metody oraz techniki. Istnieje wiele metod i technik wyceny nieruchomości, które stosuje się w zależności od stopnia rozwoju gospodarki rynkowej, a zwłaszcza od stanu rynku nieruchomości i jego obudowy prawnej oraz organizacyjnej [5].

Ze statystycznego punktu widzenia nieruchomość jest obiektem, a charakteryzujące ją atrybuty to cechy statystyczne. Do analizy rynku nieruchomości można zatem wykorzystywać różne metody statystyczne, które mogą wspomagać szacowanie rynkowej ceny

nieruchomości. W tego typu problemach, poza metodami statystycznymi, zastosowanie mają również sztuczne sieci neuronowe. Jednak rynek nieruchomości, pomimo tego że zachodzą na nim prawidłowości statystyczne, w praktyce jest trudny do badania. Problemy, które zwykle pojawiają się przy próbie wykorzystania metod statystycznych, można podzielić na trzy grupy ograniczeń [6]: prawne, metodologiczne oraz techniczne. Wpływają one między innymi na jakość, wiarygodność i liczbę danych, co utrudnia stosowanie metod statystycznych. Sztuczne sieci neuronowe mają tę przewagę nad metodami statystycznymi, że nie wymagają spełnienia żadnych założeń dotyczących np. rozkładu prawdopodobieństwa zmiennych oraz znajomości zależności zachodzących między zmiennymi opisującymi badane obiekty. Wymagają jednak wcześniejszego trenowania i wyboru odpowiedniej struktury sieci i parametrów uczenia, co nie jest łatwe w praktycznych zastosowaniach.

3. SIECI NEURONOWE JEDNOKIERUNKOWE

Sztuczna sieć neuronowa jest systemem zbudowanym z bardzo wielu połączonych, równocześnie pracujących i wspólnie przetwarzających informacje elementów, tzw. neuronów. Działanie sztucznego neuronu polega na przetworzeniu sygnałów, które otrzyma na wejściu i wygenerowaniu sygnału wyjściowego. Sygnały wejściowe są mnożone przez odpowiednie współczynniki zwane wagami, które w odniesieniu do neuronu biologicznego pełnią funkcję synaps, a więc decydują o sile połączenia. Następnie sumując „ważone” sygnały wejściowe, oblicza się (czasami dodając jeszcze wyraz wolny) całkowite pobudzenie neuronu Φ [7]:

$$\Phi = \sum_{i=1}^m w_i x_i = \mathbf{w}^T \mathbf{x}, \quad (1)$$

gdzie: \mathbf{w} – wektor współczynników wag, \mathbf{x} – wektor sygnałów wejściowych.

Następnie sygnał Φ jest przekształcany na wyjście sieci y przez odpowiednią funkcję aktywacji:

$$\mathbf{y} = \mathbf{f}(\Phi), \quad (2)$$

gdzie \mathbf{f} jest funkcją aktywacji, która może mieć różną postać. Najczęściej wykorzystuje się funkcje: progową, logistyczną, tangens hiperboliczny i inne [8].

Sieci jednokierunkowe są to sieci o jednym kierunku przepływu sygnałów – od wejścia do wyjścia. Neurony tworzą warstwy, a połączone są tylko neurony w sąsiednich warstwach [9]. Przykładem sieci jednokierunkowej jest perceptron wielowarstwowy, który składa się zawsze z warstwy wejściowej i wyjściowej, a także może posiadać jedną lub więcej warstw ukrytych. Neurony w warstwie wejściowej wstępnie przetwarzają obraz wejściowy (dokonując np. normalizacji lub skalowania sygnałów), a następnie w warstwach ukrytych i warstwie wyjściowej odbywa się zasadnicze przetwarzanie sygnałów wejściowych. Połączeniom neuronów z sąsiednich warstw odpowiadają współczynniki wag. Na wyjściu sieci generowany jest sygnał wyjściowy.

Zastosowanie sieci do rozwiązywania konkretnego zadania wymaga wcześniejszego wytrenowania sieci, wykorzystując do tego celu zbiór danych nazywany uczącym. Istnieją różne metody uczenia sieci, które stosuje się w zależności od typu sieci i zastosowań. Podstawową i najprostszą metodą uczenia sieci jednokierunkowych, tj. estymacji wartości wag, jest metoda wstecznej propagacji błędów, która jest zaliczana do grupy metod uczenia z nauczycielem. Stosuje się je, gdy znany jest pożądany sygnał wyjściowy [10]. Trenowanie sieci za pomocą algorytmu wstecznej propagacji błędów polega na przesyłaniu błędów uczenia od warstwy wyjściowej do warstwy wejściowej. Jest to algorytm oparty na minimalizacji

sumy kwadratów błędów z wykorzystaniem optymalizacyjnej metody największego spadku. Dla każdego wektora ze zbioru uczącego obliczana jest wartość wyjścia sieci, która jest następnie porównywana z wartością pożądaną. Wartości wag sieci są wyznaczone w taki sposób, aby błąd między wartością pożądaną na wyjściu a wartością obliczoną był minimalizowany. Postępowanie to jest powtarzane tak długo, aż błąd dla całego zbioru będzie dostatecznie mały lub do momentu przeprowadzenia maksymalnej liczby iteracji. W praktycznych zastosowaniach, w celu polepszenia efektów uczenia, uwzględnia się w procesie trenowania współczynnik uczenia oraz współczynnik momentum [7].

Często rozwiązując różnego rodzaju zadania za pomocą sieci neuronowych, korzysta się z innego narzędzia należącego do grupy metod sztucznej inteligencji, którym są algorytmy genetyczne. W przeprowadzonych eksperymentach klasyfikacji nieruchomości zastosowano algorytmy genetyczne do wyboru optymalnego zestawu atrybutów opisujących nieruchomości. Algorytmy te charakteryzują się uniwersalnością, elastycznością i skutecznością oraz dużą odpornością na zakłócenia występujące w danych wejściowych. Dzięki tej ostatniej właściwości algorytm pomija wszelkie czynniki, które nie mają bezpośredniego wpływu na wynik, co jest szczególnie cenne np. podczas analizowania zjawisk ekonomicznych [11].

4. OPIS I WYNIKI PRZEPROWADZONYCH BADAŃ

W celu rozwiązania zadania klasyfikacji nieruchomości do jednej z wcześniej zdefiniowanych grup cenowych, zastosowano perceptron wielowarstwowy, w którym na wejściu podawane są atrybuty nieruchomości, a na wyjściu otrzymuje się numer grupy. Dane wykorzystane w badaniach pochodziły z łódzkiego wtórnego rynku mieszkaniowego. Informacje zostały udostępnione przez agencję nieruchomości i dotyczyły transakcji przeprowadzonych za pośrednictwem tej agencji¹. Zbiór danych stanowiły mieszkania opisane przez następujące zmienne: x_1 – cena transakcyjna (w zł), x_{1*} – cena transakcyjna metra kwadratowego powierzchni (w zł), x_2 – powierzchnia mieszkania (w m^2), x_3 – numer piętra, x_4 – liczba pokoi, x_5 – rok budowy, x_6 – telefon (1 – tak, 0 – nie), x_7 – lokalizacja mieszkania (A; B; C; D; E; F; G; H; I; J; K; L; M; N; O; P; R; S; T; Z), x_8 – rozkład mieszkania (1 – rozkładowe, 0 – amfilada), x_9 – balkon (1 – jest, 0 – nie ma), x_{10} – rozmieszczenie mieszkania w budynku (1 – środkowe, 0 – szczytowe), x_{11} – ocena stanu technicznego (4; 3; 2; 1; 0). Po wstępnej analizie zbioru danych² został podzielony na cztery ustalone grupy według cen transakcyjnych poszczególnych obiektów i według cen za metr kwadratowy powierzchni (tab. 1).

Tabela 1. Struktura mieszkań według arbitralnie ustalonych grup

Grupa	I	II	III	IV
cena w tys. zł	do 50	<50-75)	<75-100)	100 i powyżej
Liczebność	54	98	59	29
Grupa	I	II	III	IV
Cena za m^2 w zł	do 1400	<1400-1600)	<1600-1800)	1800 i powyżej
Liczebność	63	87	67	23

Źródło: Opracowanie własne

Klasyfikacji obiektów dokonano na podstawie różnych zestawów zmiennych diagnostycznych. We wszystkich modelach klasyfikacyjnych wykorzystano perceptron składający się z trzech warstw: wejściowej, zawierającej tyle neuronów, ile było zmiennych wejściowych, ukrytej zawierającej trzy neurony oraz wyjściowej, zawierającej cztery

¹ Transakcje przeprowadzone w latach 2000-2001.

² W eksperymentach numerycznych przypadki z brakami informacji o wartościach zmiennych były usuwane.

neurony, odpowiadające wcześniej zdefiniowanym grupom cenowym – modele oznaczone jako: S1, S2, S3, S4 lub grupom cen za metr kwadratowy – model S2*. Wszystkie sieci trenowano metodą wstecznej propagacji błędów. Cały zbiór danych został podzielony na trzy zbiory: uczący, walidacyjny oraz testowy.

Dla sieci S1 zestaw zmiennych wejściowych wybrano, wykorzystując algorytm genetyczny i były to zmienne: $x_2, x_4, x_5, x_6, x_8, x_9, x_{10}$ i x_{11} . W kolejnych trzech modelach uwzględniono:

- wszystkie zmienne ($x_2 - x_{11}$) – model S2,
- zestaw zmiennych: x_2, x_4, x_5, x_9 oraz x_{11} – model S3,
- zestaw zmiennych: $x_2, x_5, x_6, x_7, x_9, x_{11}$ – model S4.

W przypadku klasyfikacji mieszkań do grup cen za metr kwadratowy, zmiennymi wejściowymi były wszystkie zmienne ($x_2 - x_{11}$) – sieć S2*, natomiast zmienną wyjściową była nominalna zmienna oznaczająca grupę cen za metr kwadratowy (I, II, III, IV).

Oceny wyników klasyfikacji mieszkań do ustalonych klas dokonano na podstawie wyznaczonych błędów klasyfikacji [8]:

- **ogólnego błędu klasyfikacji**, który określa procent błędnie zaklasyfikowanych obiektów w całym zbiorze:

$$E = \frac{n}{N} \cdot 100\%, \quad (3)$$

- **błędu klasyfikacji do innej klasy niż A_p** (do której dany obiekt należy), który pozwala określić, jaki procent jednostek w rzeczywistości należących do tej klasy został zaklasyfikowany do innej klasy:

$$E_p = \frac{n_p}{N_p} \cdot 100\%, \quad (4)$$

- **błędu klasyfikacji do klasy innej niż sąsiednia**, który informuje, ile jednostek z danej klasy zostało błędnie zaklasyfikowanych do klasy innej niż sąsiednia:

$$E_p^s = \frac{n_s}{N_p} \cdot 100\% \quad (5)$$

gdzie: n – liczba obiektów źle zaklasyfikowanych, n_p – liczba obiektów faktycznie należących do klasy A_p , które zostały błędnie zaklasyfikowane do innej klasy, n_s – liczba obiektów faktycznie należących do klasy A_p , które zostały błędnie zaklasyfikowane do klasy innej niż sąsiednia klasa, N – liczba wszystkich obiektów, N_p – liczba obiektów należących do klasy A_p . Jak widać w tabeli 2, dla sieci ze zmiennymi wejściowymi wybranymi przez algorytm genetyczny (S1) błędy klasyfikacji do grup innych niż sąsiednia są zerowe. W zbiorze uczącym ogólny błąd klasyfikacji wynosi 20%, a w zbiorze testowym 26,47%. Najlepiej zaklasyfikowane są obiekty z grupy II, w zbiorze testowym na dwanaście mieszkań dziesięć zostało zaklasyfikowanych poprawnie.

Tabela 2. Wyniki klasyfikacji dla sieci S1

Grupa cenowa	zbiór uczący				zbiór walidacyjny				zbiór testowy			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Razem	18	27	17	13	7	14	11	2	3	12	13	6
Poprawnie	16	25	11	8	4	11	7	1	2	10	9	4
Błędnie	2	2	6	5	3	3	4	1	1	2	4	2
Błędnie do innej niż sąsiednia klasy	0	0	0	0	0	0	0	0	0	0	0	0
E_p (%)	11,11	7,41	35,29	38,46	42,86	21,43	36,36	50,00	33,33	16,67	30,77	33,33
E_p^s (%)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
E (%)	20,00				32,35				26,47			

Źródło: Obliczenia własne

W przypadku sieci S2 (tabela 3) pojawiły się obiekty błędnie zaklasyfikowane do klasy innej niż sąsiednia. Ogólnie wyniki klasyfikacji dla tej sieci są złe. W zbiorze testowym tylko 50% obiektów zostało zaklasyfikowanych poprawnie.

Tabela 3. Wyniki klasyfikacji dla sieci S2

Grupa cenowa	zbiór uczący				zbiór walidacyjny				zbiór testowy			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Razem	17	27	20	11	5	11	11	7	6	15	10	3
Poprawnie	14	19	17	9	3	7	7	5	2	7	6	2
Błędnie	3	8	3	2	2	4	4	2	4	8	4	1
Błędnie do innej niż sąsiednia klasy	0	1	0	0	0	0	0	0	2	1	1	0
$E_p(\%)$	17,65	29,63	15,00	18,18	40,00	36,36	36,36	28,57	66,67	53,33	40,00	33,33
$E_p^s(\%)$	0,00	3,70	0,00	0,00	0,00	0,00	0,00	0,00	33,33	6,67	10,00	0,00
$E(\%)$	21,33				35,29				50,00			

Źródło: Obliczenia własne

Wyniki dla sieci S3 przedstawione są w tabeli 4. W tym przypadku w zbiorze testowym poprawnie zostało zaklasyfikowanych 80% obiektów, przy czym w grupie I błąd klasyfikacji wynosi 0%, a w grupach II i IV tylko po jednym przypadku zostało błędnie zaklasyfikowanych. Natomiast w grupie III błąd klasyfikacji wynosi aż 50%. W całym zbiorze wszystkie błędnie zaklasyfikowane obiekty znalazły się w grupach sąsiednich.

Tabela 4. Wyniki klasyfikacji dla sieci S3

Grupa cenowa	zbiór uczący				zbiór walidacyjny				zbiór testowy			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Razem	28	40	23	10	5	8	4	2	3	16	10	6
Poprawnie	21	32	8	7	3	7	4	2	3	15	5	5
Błędnie	7	8	15	3	2	1	0	0	0	1	5	1
Błędnie do innej niż sąsiednia klasy	0	0	0	0	0	0	0	0	0	0	0	0
$E_p(\%)$	25,00	20,00	65,22	30,00	40,00	12,50	0,00	0,00	0,00	6,25	50,00	16,67
$E_p^s(\%)$	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
$E(\%)$	32,67				15,79				20,00			

Źródło: Obliczenia własne

Dla sieci S4 (tabela 5) ogólny błąd klasyfikacji w zbiorze testowym wynosi 32,35%. Jedno mieszkanie z grupy II tego zbioru zostało zaklasyfikowane błędnie do grupy innej niż sąsiednia (IV). Pozostałe błędy polegały na klasyfikacji obiektów do grup sąsiednich.

Tabela 5. Wyniki klasyfikacji dla sieci S4

Grupa cenowa	zbiór uczący				zbiór walidacyjny				zbiór testowy			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Razem	18	27	17	13	7	14	11	2	3	12	13	6
Poprawnie	14	24	12	10	4	10	5	1	2	9	8	4
Błędnie	4	3	5	3	3	4	6	1	1	3	5	2
Błędnie do innej niż sąsiednia klasy	0	0	0	0	0	0	1	0	0	1	0	0
$E_p(\%)$	22,22	11,11	29,41	23,08	42,86	28,57	54,55	50,00	33,33	25,00	38,46	33,33
$E_p^s(\%)$	0,00	0,00	0,00	0,00	0,00	0,00	9,09	0,00	0,00	8,33	0,00	0,00
$E(\%)$	20,00				41,18				32,35			

Źródło: Obliczenia własne

Wyniki klasyfikacji do grup cen metra kwadratowego (sieć S2*) przedstawione są w tabeli 6. Dla tej sieci otrzymano bardzo złe wyniki, o czym świadczy bardzo wysoki błąd klasyfikacji

dla zbioru testowego (64,71%). Ponadto wysokie są też błędy w poszczególnych klasach, w których większość obiektów została zaklasyfikowana niepoprawnie. W grupie IV wszystkie obiekty zostały źle ocenione przez tę sieć.

Tabela 6. Wyniki klasyfikacji dla sieci S2*

Grupa cen metra kwadratowego pow.	Zbiór uczący				Zbiór walidacyjny				Zbiór testowy			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
Razem	14	28	23	10	7	14	11	2	8	9	13	4
Poprawnie	10	19	13	8	4	8	4	1	2	4	6	0
Błędnie	4	9	10	2	3	6	7	1	6	5	7	4
Błędnie do innej niż sąsiednia klasy	1	1	2	1	1	0	3	0	2	0	1	3
$E_p(\%)$	28,57	32,14	43,48	20,00	42,86	42,86	63,64	50,00	75,00	55,56	53,85	100,00
$E_p^s(\%)$	7,14	3,57	8,70	10,00	14,29	0,00	27,27	0,00	25,00	0,00	7,69	75,00
$E(\%)$	33,33				50,00				64,71			

Źródło: Obliczenia własne

W celu porównania wyników klasyfikacji uzyskanych przez wszystkie modele sieci neuronowych, w tabeli 7 zestawiono błędy klasyfikacji E , E_p , E_p^s w zbiorach testowych. Najlepszą jakość klasyfikacji uzyskano dla modeli S1 i S3, a najgorszą dla S2*.

Tabela 7. Błędy klasyfikacji w zbiorach testowych dla różnych modeli sieci neuronowych

Grupa	Modele sieci neuronowych									
	S1		S2		S3		S4		S2*	
	$E_p(\%)$	$E_p^s(\%)$	$E_p(\%)$	$E_p^s(\%)$	$E_p(\%)$	$E_p^s(\%)$	$E_p(\%)$	$E_p^s(\%)$	$E_p(\%)$	$E_p^s(\%)$
I	33,33	0,00	66,67	33,33	0,00	0,00	33,33	0,00	75,00	25,00
II	16,67	0,00	53,33	6,67	6,25	0,00	25,00	8,33	55,56	0,00
III	30,77	0,00	40,00	10,00	50,00	0,00	38,46	0,00	53,85	7,69
IV	33,33	0,00	33,33	0,00	16,67	0,00	33,33	0,00	100,00	75,00
$E(\%)$	26,47		50,00		20,00		32,35		64,71	

Źródło: Obliczenia własne

Na podstawie uzyskanych wyników można stwierdzić, że sieci lepiej klasyfikują do grup cenowych. Najlepszą jakość klasyfikacji (80% poprawnie zaklasyfikowanych mieszkań w zbiorze testowym, 67,33% – w zbiorze uczącym) uzyskano dla zmiennych wejściowych: x_2 , x_4 , x_5 , x_9 oraz x_{11} . Dobre wyniki klasyfikacji uzyskano również dla modelu, w którym zmienne wejściowe zostały wybrane przez algorytm genetyczny. W tym przypadku otrzymano 73,53% poprawnych klasyfikacji w zbiorze testowym, a 80% – w zbiorze uczącym.

5. PODSUMOWANIE

Zaprezentowane w artykule wyniki badań potwierdzają zasadność stosowania jednokierunkowych sieci neuronowych w analizach rynku nieruchomości. Odsetek poprawnych klasyfikacji dla najlepszego z tych modeli wynosi 80%, przy czym najgorzej klasyfikowane są przeważnie mieszkania z grupy mieszkań najdroższych, co może wynikać z faktu, że w całym zbiorze danych jest ich najmniej. Ponadto, błędy klasyfikacji dotyczą zawsze mieszkań z grup sąsiednich. Dla jakości wyników istotny okazał się wybór odpowiedniego zestawu zmiennych wejściowych. Dla pięciu zmiennych otrzymano znacznie lepsze wyniki klasyfikacji niż dla całego zbioru atrybutów. Można zatem wnioskować, że jednokierunkowe sieci neuronowe mogą być bardzo przydatne w klasyfikacji nieruchomości do ustalonych grup cenowych, pod warunkiem że w procesie tym uwzględniony będzie odpowiedni zbiór cech charakteryzujących te nieruchomości. Należy również zauważyć, że

wpływ na efektywność zastosowania sieci miał też wybór zmiennej grupującej. W przypadku gdy zmienną tą była cena za metr kwadratowy powierzchni mieszkania, nie otrzymano zadowalających wyników klasyfikacji.

Weryfikując efektywność zastosowanych metod, należy wziąć pod uwagę fakt, iż uzyskanie dobrej jakości klasyfikacji wymaga przede wszystkim posiadania dostatecznie licznej i kompletnego zbioru informacji o nieruchomościach. W artykule przedstawiono wyniki eksperymentów opartych na stosunkowo niewielkim zbiorze danych empirycznych (co wynika z trudnej ich dostępności), który ponadto zawierał dużo obiektów z brakami informacji o wartościach niektórych zmiennych, co może mieć wpływ na pogorszenie wyników. Poza niewielką liczebnością zbioru danych, na jakość klasyfikacji ma wpływ również stosunkowo mała liczba obiektów w niektórych grupach (największe błędy otrzymano dla grupy najmniej licznej). Należy również zauważyć, że przeprowadzone badania dotyczyły klasyfikacji mieszkań do ustalonych arbitralnie czterech grup cenowych, a – jak to wynika z badań empirycznych – w przypadku klasyfikacji do więcej niż dwóch klas różnice między nimi zazwyczaj się zacierają. To właśnie może być przyczyną zawyżenia ogólnych błędów klasyfikacji przy obserwacji, że w większości eksperymentów błędne rozpoznawanie mieszkań polegało na klasyfikowaniu ich do klasy sąsiedniej.

LITERATURA

- [1] Mulawka J.J.: *Systemy ekspertowe*, Wydawnictwo Naukowo-Techniczne, Warszawa, 1996, s. 137.
- [2] Kucharska-Stasiak E.: *Nieruchomość a rynek*, Wydawnictwo Naukowe PWN, Warszawa, 1999, s. 30, 19-24.
- [3] Bryx M., Matkowski R.: *Inwestycje w nieruchomości*, Poltext, Warszawa, 2001, s. 70-76.
- [4] Śliwiński A.: *Zarządzanie nieruchomościami. Podstawy wiedzy i praktyki gospodarowania nieruchomościami*, Agencja Wydawnicza „Placet”, Warszawa, 2000, s. 45-46.
- [5] Hopfer A.(red.): *Wycena nieruchomości*, Wydawnictwo ART, Olsztyn, 1991, s. 8.
- [6] Foryś I., Kokot S.: *Problemy badania rynku nieruchomości*, Zeszyty Naukowe Uniwersytetu Szczecińskiego, nr 318, Prace Katedry Ekonometrii i Statystyki nr 10, 2001, s. 175-182.
- [7] Korbicz J., Obuchowicz A., Uciński D.: *Sztuczne Sieci Neuronowe. Podstawy i zastosowania*, Akademicka Oficyna Wydawnicza PLJ, Warszawa, 1994, s. 23, 50-58.
- [8] Witkowska D.: *Sztuczne sieci neuronowe i metody statystyczne. Wybrane zagadnienia finansowe*, C.H. BECK, Warszawa, 2002, s. 4.
- [9] Osowski S.: *Sieci neuronowe*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 1994, s. 19.
- [10] Tadeusiewicz R.: *Wprowadzenie do praktyki stosowania sieci neuronowych*, [w:] *Sieci neuronowe*, StatSoft, Kraków, 1999, s. 19.
- [11] Gwiazda T.D.: *Algorytmy genetyczne. Zastosowanie w finansach*, Wyższa Szkoła Przedsiębiorczości i Zarządzania im. L. Koźmińskiego, Warszawa, 1998, s. 50-54.